REMARKS

The application is believed to be in condition for allowance because the claims are novel and non-obvious over the cited art. The following paragraphs provide the justification for these beliefs. In view of the following reasoning for allowance, the applicant hereby respectfully requests further examination and reconsideration of the subject application.

Claims 1-32 are pending in this application.

Objection to the Specification

The Abstract was objected to for having the title of the application on the Abstract page and being longer than 150 words. The applicants have amended the specification to overcome this objection by removing the title and shortening the Abstract.

The Section 101 Rejection of Claims 1-22 and 23-32

Claims 1-22 and 23-32 were rejected under 35 USC 101 because the claimed invention is allegedly directed to non-statutory subject matter. The applicants respectfully traverse this contention that the claims are directed towards non-statutory subject matter.

Specifically, Claim 1 was amended to specify that the claimed system uses a computer to perform the claimed process actions. It should be noted that the applicant is not claiming a computer program *per se*. The preamble of independent Claim 1 now reads,

> "A system for automatically decomposing an image sequence, comprising using a computer to perform the following process actions:"

Thus, the applicant is claiming a process implemented on a computer where the actions of the process are performed using the computer. This is statutory subject matter. As stated in the MPEP (see Section 2106 (IV)(B)(1)(a) at Page 2100-13, Rev. 2, May 2004):

> "Computer programs are often recited as part of a claim. Office personnel should determine whether the computer program is being claimed as part of an otherwise statutory manufacture or machine. In such a case, the claim remains statutory irrespective of the fact that a computer program is included in the claim. **The same result occurs when a computer program is used in a computerized process where the computer executes the instructions set forth in the computer program.**"

As to Claims 23-32, the MPEP Section 2106.01 states,

> "When a computer program is claimed in a process where the computer is executing the computer program's instructions, USPTO personnel should treat the claim as a process claim."

Hence, Claims 23-32 are statutory subject matter per the MPEP Section 2106.01.

Furthermore, while the applicant does not admit to and does not believe that a modulated signal is non-statutory subject matter, it has been decided to amend the specification to eliminate any reference to a computer readable medium including a modulated signal such as a carrier wave. The intent of the amendment to the specification is to limit the claimed invention to the use of *physical* computer readable media, and that the use of carrier waves is not intended to be included in the scope of the claimed invention.

In view of the amended specification and claims, it is believed that claims 1-32 are patentable under 35 USC 101. Therefore, it is respectfully requested that the rejection of these claims be reconsidered.

<u>The Rejection of Claims 1-3, 5-6, 14, 18-19 and 23-24 Under 35 USC 102(b).</u>

Claims 1-3, 5-6, 14, 18-19 and 23-34 stand rejected under 35 USC 102(b) as being anticipated by Foote et al. U.S. Patent No. 6,404,925 (hereinafter Foote). It was contended in the above-identified Office Action that Foote teaches all the elements of the rejected claims. The applicants respectfully disagree with this contention of anticipation.

The applicants claim a technique that can extract objects from an image sequence using the constraints on their motion and also performs tracking while the appearance models are learned. **The technique operates in near real time, processing data and learning generative models at substantially the same rate the input data is received. (Summary)**

The claimed technique tries to recognize patterns in time (e.g., finding possibly recurring scenes or objects in an image sequence), and in order to do so attempts to model the process that could have generated the pattern. It uses the possible states or classes, the probability of each of the classes being in each of the states at a given time and a state transition matrix that gives the probability of a given state given that state at a previous time. The states further may include observable states and hidden states. In such cases the observed sequence of states is probabilistically related to the hidden process. The processes are modeled using a transformed Hidden Markov model (THHM) where there is an underlying hidden Markov process changing over time, and a set of observable states which are related somehow to the hidden states. The connections between the hidden states and the observable states represent the probability of generating a particular observed state given that the Markov process is in a particular hidden state. All probabilities entering an observable state will sum to 1. (Summary)

The number of classes of objects and an image sequence is all that must be provided in order to extract objects from an image sequence and learn their generative model (e.g., a model of how the observed data could have been generated). **Given this information, probabilistic inference and learning are**

used to compute a single set of model parameters that represent either the video sequence processed to that point or the entire video sequence. These model parameters include the mean appearance and variance of each class. The probability of each class is also determined. (Summary)

More specifically, the applicants claim,

> "A system for automatically decomposing an image sequence, comprising using a computer to perform the following process actions:
>
> providing an **image sequence of at least one image frame of a scene**;
>
> providing **a preferred number of classes of objects to be identified within the image sequence**;
>
> **automatically decomposing the image sequence into the preferred number of classes of objects in near real-time.**"

And,

> "A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising using a computer to:
>
> **acquire at least one image sequence, each image sequence having at least one image frame;**
>
> in near real time **automatically decompose each image sequence into a generative model, with each generative model including a set of model parameters that represent at least one object class for each image sequence** using an expectation-maximization analysis that employs a Viterbi analysis. "

Foote discloses methods for segmenting audio-video recording of meetings containing slide presentations by one or more speakers. These segments serve as indexes into the recorded meeting. If an agenda is provided for the meeting, these segments can be labeled using information from the agenda. The system automatically detects intervals of video that correspond to presentation slides. **Under the assumption that only one person is speaking during an interval when slides are displayed in the video, possible speaker intervals are extracted from the audio soundtrack by finding these regions.** Since the same speaker may talk across multiple slide intervals, **the acoustic data from these**

intervals is clustered to yield an estimate of the number of distinct speakers and their order. Clustering the audio data from these intervals yields an estimate of the number of different speakers and their order. Merged clustered audio intervals corresponding to a single speaker are then used as training data for a speaker segmentation system. Using speaker identification techniques, **<u>the full video is then segmented into individual presentations based on the extent of each presenter's speech</u>**. (Abstract)

Foote does not teach the applicants' claimed preferred <u>number</u> of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects <u>in near real-time</u>. Nor does Foote teach <u>in near-real time</u> automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis.

Granted, as to Claim 1, the Office Action states that providing an image sequence of at least one image frame is taught in FIG. 2, element 201 and FIG. 3, elements 301-308. But FIG. 3 refers to training images for training the Foote system shown in FIG. 2, not an image frame of element 201. Additionally, the Office Action states that providing a preferred number of classes of objects is taught as a "pre-defined set of classes" in Col. 5, lines 14-16 to be identified within the image sequence. **But a "predefined set of classes" is not the same as a preferred <u>number</u> of classes, as the applicants claim.** In the applicants' claimed invention it is not necessary to define what type of a class is sought, all that is needed is the preferred <u>number</u> of classes sought, which requires much less information to specify than a class itself. Furthermore, Claim 1 includes the limitation of automatically decomposing the image sequence into the preferred number of classes of objects **<u>in near real-time</u>**." Cited Column 5, lines 14-16, does not teach "automatically decomposing the image sequence into the preferred number of classes of objects **<u>in</u>**

near real-time". Nothing at all is stated in this paragraph regarding processing in near real-time. In fact, clearly Foote does not teach automatically decomposing the image sequence into **the preferred number** of classes of objects **in near real-time** because Foote segments a **full video into individual presentations based on the extent of each presenter's speech**. (Abstract) Hence, Foote can only segment a video file with corresponding audio after it has been recorded, not in real-time as it is being input.

As for Claim 23, the Office Action states that providing an image sequence of at least one image frame is taught in FIG. 2, element 201 and FIG. 3, elements 301-308. But FIG. 3 refers to training images for training the Foote system shown in FIG. 2, not an image frame of element 201. Furthermore, the Office Action states that automatically decomposing each image sequence into a generative model is taught in FIG. 2, elements 202-205; Col. 5, line 65- Col. 6 line 2, but this passage does not teach automatically decomposing each image sequence into a generative model. It merely appears to determine video features in image frames and using these features to determine which of the predefined classes a frame belongs to. It does not teach automatically decomposing each image sequence into a generative model (e.g., a model of how the observed data could have been generated) with each generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Finally, nothing in Foote teaches the decomposing of an image sequence in near real-time.

Thus, the applicants have claimed an element not taught in Foote, namely inputting a number of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects in near real-time. Also Foote does not teach decomposing an image sequence into a generative model or decomposition of an image sequence in near real time. As such, the rejected claims, as amended, are not anticipated by the reference. It is, therefore, respectfully requested that the rejection of Claims 1-3, 5-

6, 14, 18-19 and 23-34 be reconsidered based on the distinguishing claim language, i.e.:

> "A system for automatically decomposing an image sequence, comprising using a computer to perform the following process actions:
>
> providing an **image sequence of at least one image frame of a scene**;
>
> providing **a preferred number of classes of objects to be identified within the image sequence**;
>
> **automatically decomposing the image sequence into the preferred number of classes of objects in near real-time.**"

And,

> "A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising using a computer to:
>
> **acquire at least one image sequence, each image sequence having at least one image frame;**
>
> **in near real-time automatically decompose each image sequence into a generative model, with each generative model including a set of model parameters that represent at least one object class for each image sequence** using an expectation-maximization analysis that employs a Viterbi analysis. "

The 35 USC 103(a) Rejection of Claims 4, 7 and 27.

Claims 4, 7 and 27 were rejected under 35 USC 103(a) as unpatentable over Foote, in view of Petrovic et al (Transformed Hidden Markov Models: Estimating Mixture Models of Images and Inferring Spatial Transformations in Video Sequences, Computer Visions and Pattern Recognition, 2000, Vol. 2, pg 16-33), hereinafter Petrovic. The Office Action contended that Foote teaches all of the limitations of Claims 4, 7 and 27, except that Foote does not teach a model that employs a latent image and a translation variable in learning each object class, nor does Foote teach using a latent image and a translation variable in filling in hidden variables. However, the Office Action contended that Petrovic teaches these features, rendering Claims 4, 7 and 27 obvious. The applicants respectfully disagree with this contention of obviousness.

In order to deem the applicant's claimed invention unpatentable under 35 USC 103, a prima facie showing of obviousness must be made. To make a prima facie showing of obviousness, all of the claimed elements of an applicant's invention must be considered, especially when they are missing from the prior art. If a claimed element is not taught in the prior art and has advantages not appreciated by the prior art, then no prima facie case of obviousness exists. The Federal Circuit court has stated that it was error not to distinguish claims over a combination of prior art references where a material limitation in the claimed system and its purpose was not taught therein (*In Re Fine*, 837 F.2d 107, 5 USPQ2d 1596 (Fed. Cir. 1988)).

As discussed above, the applicants claim,

> "A system for automatically decomposing an image sequence, comprising using a computer to perform the following process actions:
>
> providing an **image sequence of at least one image frame of a scene**;
>
> providing **<u>a preferred number of classes of objects to be identified within the image sequence</u>**;
>
> **automatically decomposing the image sequence into the preferred number of classes of objects in near real-time."**

And,

> "A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising using a computer to:
>
> **acquire at least one image sequence, each image sequence having at least one image frame**;
>
> <u>in near real time</u> **automatically decompose each image sequence into a generative model, with each generative model including a set of model parameters that represent at least one object class for each image sequence** using an expectation-maximization analysis that employs a Viterbi analysis. "

As discussed above **Foote does not teach the applicants' claimed preferred <u>number</u> of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects <u>in near real-time</u>. Nor does Foote teach <u>in near-real time</u> automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least**

one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Petrovic also does not teach these features.

Accordingly, Foote in combination with Petrovic does not teach the applicant's claimed preferred number of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects in near real-time. Nor does Foote teach in near-real time automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Nor does Foote in combination with Petrovic recognize the advantages of the applicants' claimed invention. Namely, Foote in combination with Petrovic does not teach allowing video sequences to be decomposed into a preferred number of classes in real-time. Thus, the applicants have claimed elements not taught in the cited art and which have advantages not recognized therein. Accordingly, no prima facie case of obviousness has been established in accordance with the holding of *In Re Fine*. This lack of prima facie showing of obviousness means that the rejected claims are patentable under 35 USC 103 over Foote in view of Petrovic. As such, it is respectfully requested that Claims 4, 7 and 27 be allowed based on the previously-quoted claim language.

The 35 USC 103(a) Rejection of Claims 8-10, 13, 15-17 and 28-31.

Claims 8-10, 13, 15-17 and 28-31 were rejected under 35 USC 103(a) as unpatentable over Foote in view of Dellaert (The Expectation Maximization Algorithm, College of Computing, Georgia Institute of Technology, Technical Report Number GIT-GVU-02-20, 2/2002), hereinafter referred to as Dellaert. The Office Action contended that Foote teaches all of the limitations of Claims 8-10, 13, 15-17 and 28-31, except that Foote does not directly teach various computations in the expectation step of the generalized expectation-maximization parameters. However, the Office Action contended that Dellaert teaches these features, rendering Claims 8-10, 13, 15-17 and

28-31 obvious. The applicants respectfully disagree with this contention of obviousness.

As discussed above, the applicants claim,

> "A system for automatically decomposing an image sequence, comprising using a computer to perform the following process actions:
>
> providing an **image sequence of at least one image frame of a scene**;
>
> providing **a preferred number of classes of objects to be identified within the image sequence**;
>
> **automatically decomposing the image sequence into the preferred number of classes of objects in near real-time.**"

And,

> "A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising using a computer to:
>
> **acquire at least one image sequence, each image sequence having at least one image frame;**
>
> in near real time **automatically decompose each image sequence into a generative model, with each generative model including a set of model parameters that represent at least one object class for each image sequence** using an expectation-maximization analysis that employs a Viterbi analysis. "

As discussed above **Foote does not teach the applicants' claimed preferred number of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects in near real-time. Nor does Foote teach in near-real time automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Dellaert also does not teach these features.**

Accordingly, Foote in combination with Dellaert does not teach the applicant's claimed preferred number of classes of objects to be identified within the image

sequence or automatically decomposing the image sequence into the preferred number of classes of objects in near real-time. Nor does Foote teach in near-real time automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Nor does Foote in combination with Dellaert recognize the advantages of the applicants' claimed invention. Namely, Foote in combination with Dellaert does not teach allowing video sequences to be decomposed into a preferred number of classes in real-time. Thus, the applicants have claimed elements not taught in the cited art and which have advantages not recognized therein. Accordingly, no prima facie case of obviousness has been established in accordance with the holding of *In Re Fine*. This lack of prima facie showing of obviousness means that the rejected claims are patentable under 35 USC 103 over Foote in view of Dellaert. As such, it is respectfully requested that Claims 8-10, 13, 15-17 and 28-31 be allowed based on the previously-quoted claim language.

The 35 USC 103(a) Rejection of Claims 11-12.

Claims 11-12 were rejected under 35 USC 103(a) as unpatentable over Foote, in view of Dellaert, in further view of Eberman et al., U.S. Patent No. 5,925,065, herein after Eberman. The Office Action contended that Foote and Dellaert teach all of the limitations of Claims 11-12, except that Foote and Dellaert do not directly teach accelerating the expectation step using a FFT-based inference analysis. However, the Office Action contended that Eberman teaches this feature, rendering Claims 11-12 obvious. The applicants respectfully disagree with this contention of obviousness.

As discussed above, the applicants claim,

"A system for automatically decomposing an image sequence, comprising using a computer to perform the following process actions:

providing an **image sequence of at least one image frame of a scene**;

providing **a preferred number of classes of objects to be identified within the image sequence**;

automatically decomposing the image sequence into the preferred number of classes of objects in near real-time."

As discussed above **Foote does not teach the applicants' claimed preferred <u>number</u> of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects <u>in near real-time</u>. <u>Dellaert and Eberman also do not teach these features.</u>**

Accordingly, Foote in combination with Dellaert and Eberman do not teach the applicant's claimed preferred <u>number</u> of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects <u>in near real-time</u>. Nor does Foote in combination with Dellaert and Eberman recognize the advantages of the applicants' claimed invention. Namely, Foote in combination with Dellaert and Eberman does not teach allowing video sequences to be decomposed into a preferred number of classes in real-time. Thus, the applicants have claimed elements not taught in the cited art and which have advantages not recognized therein. Accordingly, no prima facie case of obviousness has been established in accordance with the holding of *In Re Fine*. This lack of prima facie showing of obviousness means that the rejected claims are patentable under 35 USC 103 over Foote in view of Dellaert. As such, it is respectfully requested that Claims 11-12 be allowed based on the previously-quoted claim language.

<u>The 35 USC 103(a) Rejection of Claims 20-21 and 25-26.</u>

Claims 20-21 and 25-26 were rejected under 35 USC 103(a) as unpatentable over Foote, in view of Jojic et al (Learning Flexible Sprites in Video Layers, Proc. Of IEEE Conf. on Computer Vision and Pattern Recognition, 2001, pg. 1-8). The Office Action contended that Foote teaches all of the limitations of claims, except that Foote does not various model parameters of the applicants' claimed invention. However, the Office Action contended that Jojic teaches these features, rendering Claims 20-21 and

25-26 obvious. The applicants respectfully disagree with this contention of obviousness.

As discussed above, the applicants claim,

> "A system for automatically decomposing an image sequence, comprising using a computer to perform the following process actions:
>
> providing an **image sequence of at least one image frame of a scene**;
>
> providing **a preferred number of classes of objects to be identified within the image sequence**;
>
> **automatically decomposing the image sequence into the preferred number of classes of objects in near real-time.**"

And,

> "A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising using a computer to:
>
> **acquire at least one image sequence, each image sequence having at least one image frame**;
>
> in near real time **automatically decompose each image sequence into a generative model, with each generative model including a set of model parameters that represent at least one object class for each image sequence** using an expectation-maximization analysis that employs a Viterbi analysis. "

As discussed above **Foote does not teach the applicants' claimed preferred number of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects in near real-time. Nor does Foote teach in near-real time automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Jojic also does not teach these features.**

Accordingly, Foote in combination with Jojic does not teach the applicant's claimed preferred number of classes of objects to be identified within the image

sequence or automatically decomposing the image sequence into the preferred number of classes of objects in near real-time. Nor does Foote in combination with Jojic teach in near-real time automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Nor does Foote in combination with Jojic recognize the advantages of the applicants' claimed invention. Namely, Foote in combination with Jojic does not teach allowing video sequences to be decomposed into a preferred number of classes in real-time. Thus, the applicants have claimed elements not taught in the cited art and which have advantages not recognized therein. Accordingly, no prima facie case of obviousness has been established in accordance with the holding of *In Re Fine*. This lack of prima facie showing of obviousness means that the rejected claims are patentable under 35 USC 103 over Foote in view of Petrovic. As such, it is respectfully requested that Claims 20-21 and 25-26 be allowed based on the previously-quoted claim language.

The 35 USC 103(a) Rejection of Claim 32.

Claim 32 was rejected under 35 USC 103(a) as unpatentable over Foote, in view Eberman. The Office Action contended that Foote and Eberman teach all of the limitations of Claim 11 which recites identical features to Claim 32, and Claim 32 is thus obvious with the reason as previously described for Claim 11. The applicants respectfully disagree with this contention of obviousness.

As discussed above, the applicants claim,

> "A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising using a computer to:
>
> **acquire at least one image sequence, each image sequence having at least one image frame;**
>
> **in near real time automatically decompose each image sequence into a generative model, with each generative model including a set of model parameters that represent at least one object class for each**

image sequence using an expectation-maximization analysis that employs a Viterbi analysis. "

As discussed above **Foote does not teach the applicants' claimed preferred <u>number</u> of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects <u>in near real-time</u>. Nor does Foote teach <u>in near-real time</u> automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. <u>Eberman also does not teach these features.</u>**

Accordingly, Foote in combination with Eberman does not teach the applicant's claimed preferred <u>number</u> of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects <u>in near real-time</u>. Nor does Foote in combination with Eberman teach <u>in near-real time</u> automatically decomposing each image sequence into a generative model including a set of model parameters that represent at least one object class for each image sequence using an expectation-maximization analysis that employs a Viterbi analysis. Nor does Foote in combination with Eberman recognize the advantages of the applicants' claimed invention. Namely, Foote in combination with Eberman does not teach allowing video sequences to be decomposed into a preferred number of classes in real-time. Thus, the applicants have claimed elements not taught in the cited art and which have advantages not recognized therein. Accordingly, no prima facie case of obviousness has been established in accordance with the holding of *In Re Fine*. This lack of prima facie showing of obviousness means that the rejected claims are patentable under 35 USC 103 over Foote in view of Eberman. As such, it is respectfully requested that Claim 32 be allowed based on the previously-quoted claim language.

The applicants hereby respectfully request reconsideration of the subject application and allowance of Claims 1-32 at an early date.

LYON & HARR, LLP
300 Esplanade Drive
Suite 800
Oxnard, CA 93036
(805) 278-8855

Respectfully submitted,

Katrina A. Lyon

Katrina A. Lyon
Reg. No. 42,821
Attorney for Applicant(s)